UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS

SINGULAR COMPUTING LLC, Plaintiff, v. GOOGLE LLC, Defendant.	Civil Action No. 1:19-cv-12551 FDS Hon. F. Dennis Saylor IV

DECLARATION OF GU-YEON WEI

I. INTRODUCTION

1. I am the Robert and Suzanne Case Professor of Electrical Engineering and Computer Science in the John A. Paulson School of Engineering and Applied Sciences at Harvard University.

2. I have been asked by counsel for Google LLC to render an opinion with respect to the claim term "repeated execution" in U.S. Patent Nos. 8,407,273; 9,218,156; and 10,416,961 (collectively, the "asserted patents").

3. I am being compensated at the rate of $900 per hour for preparation of expert reports, research, and related tasks, and deposition testimony. My fee is not contingent on any ruling or outcome in the litigation.

4. In arriving at my opinions, I have relied on my education, experience, and knowledge gained over more than 20 years as an engineer and researcher. The opinions presented in this declaration are my own and, to the best of my ability, are true and correct.

II. BASIS FOR OPINIONS

A. Qualifications

5. As noted above, I am the Robert and Suzanne Case Professor of Electrical Engineering and Computer Science in the John A. Paulson School of Engineering and Applied Sciences at Harvard University. I also am a Fellow at Samsung Research in South Korea.

6. I teach courses in electronic devices and circuits, VLSI design, and mixed-signal circuits and systems. I also supervise experimental research in digital, analog, and mixed-signal integrated circuits (ICs) for low-power, energy-efficient computing systems and architectures, which include circuit design, layout, and testing of numerous analog, digital, and mixed-signal

chips. For the past six years, my hardware research has focused increasingly on accelerators for machine learning and use of machine learning for hardware design.

7. I received my BS, MS, and PhD in Electrical Engineering from Stanford University in 1994, 1997, and 2001, respectively. My coursework includes undergraduate- and graduate-level courses in circuit design, semiconductor devices, and semiconductor process technology. My PhD dissertation focused on energy-efficient parallel interface circuits that utilize adaptive power-supply regulation. This work includes the IC design and layout of on-chip pulse-width modulated switching regulators. Prior to joining the faculty at Harvard, I was a Senior Design Engineer at Accelerant Networks (acquired by Synopsys in 2003), where I designed clocking circuitry for a 5-Gbps backplane transceiver. From 2002 to 2004, I was a Design Engineering Consultant at Analog Devices in Woburn, MA. My experience also includes summer internships at Hyundai Electronics (now Hynix), Silicon Graphics, and Texas Instruments. Additional details of my career can be found my Curriculum Vitae, which is attached as **Exhibit A**.

8. I have numerous publications in several conferences and journals in the field of integrated circuits, computer architecture, and design automation, all focused towards computing for machine learning and leveraging machine learning for efficient computing. They include the International Symposium on Low-Power Electronics and Design (ISLPED), the International Solid-State Circuits Conference (ISSCC), the International Symposium on Circuits and Systems (ISCAS), the Custom Integrated Circuits Conference (CICC), the Symposium on VLSI Circuits (SoVC), the International Symposium on Quality of Electronic Design (ISQED), the Design Automation and Test in Europe Conference (DATE), the Design Automation Conference (DAC), the International Conference on Computer Aided Design (ICCAD), the Asian Solid-

State Circuits Conference (A-SSCC), the International Symposium on Code Generation and Optimization (CGO), the Energy Conversion Congress and Exposition (ECCE), the International Conference on Intelligent Robots and Systems (IROS), the International Conference on Robotics and Automation (ICRA), the International Symposium on Computer Architecture (ISCA), the High-Performance Computer Architecture (HPCA), the Symposium on Microarchitecture (micro), ASPLOS, the IEEE international Symposium on Workload Characterization (IISWC), the Conference on Machine Learning and Systems (MLSys), the International Conference on Machine Learning (ICML), the Journal of Solid-State Circuits (JSSC), the Transactions on Architecture and Code Optimization (TACO), the Journal on Low Power Electronics (JOLPE), the Electronic Letters, the Sensors and Actuators A: Physical, the IEEE micro (MICRO), the Transactions on Circuits and Systems, II (TCAS-II), and others. I have also served as reviewer for papers submitted to numerous conferences and journals. I am (or have been) on the technical program committee for numerous conferences such as ISLPED, ISCAS, and the International Conference on Computer-Aided Design (ICCAD), and have served as Session Chair several times. I was an Associate Guest Editor for a special issue of TCAS-II in 2003 and served as an Associate Editor for TCAS-II. A list of publications authored by me in the previous 20 years can also be found in my CV.

9. Last, my CV includes a list of cases in which I have been deposed and/or testified at trial or a hearing.

B. Legal Understandings

10. I am not a lawyer and do not have formal training in the interpretation of patent claims. I understand that claim construction is a matter of law and that the Court has the ultimate

responsibility to construe the claims and, in exercising that responsibility, may find it useful to consider the perspective of a person having knowledge and experience in the relevant art.

11. In providing my perspective, I understand that terms used in patent claims are generally given the meaning that the term would have to a person of ordinary skill in the art at the time of the invention. I also understand that the person of ordinary skill in the art is deemed to read the claim term not only in the context of the particular claim in which the disputed term appears, but in the context of the entire patent, including the specification. I also understand that a claim term is indefinite if it fails to inform a person of ordinary skill in the art as to what is claimed with reasonable certainty.

12. In my opinion, a person of ordinary skill in the art related to the asserted patents would have had at least a bachelor's degree in Electrical Engineering, Computer Engineering, Applied Mathematics, or the equivalent, and at least two years of academic or industry experience in computer architecture. More education could substitute for experience, and vice versa.

III. TECHNOLOGY BACKGROUND

13. The asserted patents claim a device comprising at least one "low precision high-dynamic range (LPHDR) execution unit." *E.g.*, U.S. Patent No. 8,407,273 ('273 patent) at 29:65-67.[1] The claimed LPHDR execution unit is adapted to execute an operation (*e.g.*, an arithmetic operation like multiplication) on an input signal representing a first numerical value to produce an output signal representing a second numerical value. '273 patent at 29:65-30:15.

[1] Because all three asserted patents share a common specification and materially identical claim language, I cite only the '273 patent.

14. As relevant to this opinion, the claimed LPHDR execution unit must execute arithmetic operations in a manner that produces at least some results that differ from a mathematically exact result of the same operation. To give a simple example of such an inexact result: a unit executing the operation "2 x 1" could yield a result of 2.1 instead of 2.

15. The claimed difference in results is not, however, measured according to the result generated by each separate instance of executing the operation on a given input. Instead, the asserted patents claim that, for a certain percentage (X) of possible inputs, the "repeated execution" of an operation will yield resulting "numerical values" whose "statistical mean" differs by at least a certain percentage (Y) from "the result of an exact mathematical calculation" of the same operation. Put more succinctly, the claims require an LPHDR execution unit that, if it were to perform a certain operation multiple times on the same input(s), would yield an average result that differs from an exact result by at least Y%, for at least X% of possible valid inputs.

16. For example, independent claim 1 of the '273 patent states that

> for at least X=5% of the possible valid inputs to the first operation, the statistical mean, over repeated execution of the first operation on each specific input from the at least X % of the possible valid inputs to the first operation, of the numerical values represented by the first output signal of the LPHDR unit executing the first operation on that input differs by at least Y=0.05% from the result of an exact mathematical calculation of the first operation on the numerical values of that same input.

'273 patent at 30:6-15.

17. The above claim language is identical in all the asserted patents' independent claims, except that they may vary with respect to the "X" percentages of possible valid inputs and/or "Y" percentages of difference from exact results.

18. The asserted patents make clear that the results produced by the LPHDR execution units can be "deterministic ***or*** non-deterministic." '273 patent at 27:40 (emphasis

added). A deterministic system is one that involves no randomness in the calculation of an output from an input. So, in a deterministic system, executing the same operation on the same input will always yield the same output. Conversely, a non-deterministic system may produce different results even when executing the same operation on the same input more than once.

19. The difference between deterministic and non-deterministic systems is a critical distinction between two broad types of embodiments described by the asserted patents: (1) digital and (2) analog or hybrid analog-digital. *E.g.*, '273 patent at 11:29-39, 14:16-21.

20. In computer science, a digital signal is a signal that represents data as one of a finite number of discrete values. Such signals are typically binary—that is, they have only two possible values. These values are represented by voltage bands, one of which is near a reference value (typically no volts) and one of which is near the supply voltage (the voltage level obtained by the circuit from its power supply). The voltage bands correspond to the values "zero" and "one" (or "false" and "true"). As the asserted patents teach, digital circuits are "designed to attempt to switch the transistors between completely on and completely off states." '273 patent at 6:41-43. Each value in a binary system thus represents one bit of data, which can be combined with other bits to represent larger values.

21. An analog signal, by contrast, represents data as continuous values that correspond to real numbers within a given range. For instance, in an analog system, numerical values can be represented as "charges, currents, voltages, frequencies, pulse widths, pulse densities, [or] various forms of spikes." '273 patent at 14:16-21. Variances in these signals represent different numerical values.

22. Digital systems are typically deterministic; that is, they yield the same result every time they execute the same operation on the same input. This determinism results from the

computing hardware operating on discrete representations of values, which are inherently robust and not sensitive to "noise" that affects analog systems (a phenomenon discussed below).

23. To illustrate how deterministic operations work, consider our earlier example of the operation "2 x 1," as executed by a conventional digital system. If the system yielded exact results when executing the operation, the system would always yield the result of 2. If the system executed operations that yielded inexact results, it would always yield another number, such as 2.1. Either way, when the operation is repeated on the same input, the result is always the same.

24. Although the asserted patents allege that some digital embodiments "may not yield deterministic . . . results," '273 patent at 26:43-45, it was well-known when the underlying applications were filed that a typical digital system is deterministic and, therefore, yielded predictable results when executing a given operation on a particular input. An example of a non-deterministic digital system could be one that includes an element that purposefully introduces randomness into the system's operations, such as a random number generator. In such a case, the digital system would yield non-deterministic results due to a deliberate design choice and would share the features of analog and hybrid analog-digital systems that are described below.

25. In contrast to a conventional digital system, analog and hybrid analog-digital systems are non-deterministic because of a phenomenon known as "noise." "Noise" in this context refers to a disturbance in an electrical signal, which can cause the signal's quality to deviate from the original/intended value. This signal deviation, in turn, prevents the system from generating repeatable results when executing a given operation on a particular input. For example, in an analog system, one execution of the operation "2 x 1" could yield a result of 1.9, while the next execution of the same operation could yield a result of 2.1.

26. A large number of physical conditions can introduce noise into analog operations. These conditions can exist within the analog circuit itself or be byproducts of the external environment. Examples of noise in circuits include thermal noise, 1/*f* noise, shot noise, power supply noise, and inter-signal coupling.

27. While noise can interfere with digital systems as well as analog ones, because digital signals represent discrete values, a high level of noise would be required to affect signal integrity. In fact, such high levels of noise would likely compromise the integrity of the entire system as opposed to only causing non-deterministic variations in computed results. Analog systems, by contrast, are more susceptible to noise because their signals represent continuous values, not just the value 0 or 1 (as in digital systems).

28. As a result of noise, the asserted patents' analog embodiments could not generate predictable, repeatable results when executing any given operation. The same problem of non-repeatability would affect hybrid analog-digital embodiments that "represent[] values as a mixture of analog and digital forms." '273 patent at 14:53-54. Like a purely analog system, the analog components of a hybrid system could be susceptible to noise and thus would yield results generated by the system as a whole that are not repeatable.

29. Although the effects of noise can be mitigated in analog systems, those effects cannot be eliminated. Hence, the results of the execution of analog or hybrid analog-digital operation are not repeatable. Even if the degree of variability in results can be reduced, there will always be some variability.

30. As a result, in an analog or hybrid analog-digital system, the ***average*** result of a particular operation will be prone to change with each additional execution of the operation over which the results are averaged. By extension, the degree of difference between the average result

of the operation and the mathematically exact result of the same operation will be prone to change with each additional execution of the operation.

31. As an illustration, consider again the example of performing the equation "2 x 1," but now in an analog system. The first execution of the operation could yield a result of 1.9, and a second execution could yield a result of 2.1. After the first execution, the average result (1.9 ÷ 1) would be 1.9, which differs from an exact result by 5%. But after the second execution, the average result ((1.9 + 2.1) ÷ 2) would be 2, which is the same as an exact result. Each new execution could change the average of the results of all the executions—and by extension, the difference between the average result and exact result—again.

IV. OPINION

32. In my opinion, the claim term "repeated execution" in the asserted patents' independent claims is indefinite because the patents' claims, specification, and prosecution history do not inform one of ordinary skill with reasonable certainty how many repeated executions of an operation are considered in calculating the "statistical mean" of resulting values to determine whether that statistical mean differs from an exact result by the claimed percentage.

33. To recap: the LPHDR execution units perform operations with results that differ from an exact calculation, but that difference is not measured on an execution-by-execution basis. Instead, whether the execution units meet the difference level in a given claim turns on whether the units' "repeated execution" of an operation on a certain percentage of inputs will yield results in the form of "numerical values" whose "statistical mean" differs by a specified percentage from the "exact mathematical calculation" of the same operation. *E.g.*, '273 patent at 30:8-15.

34. In other words, the asserted patents contemplate evaluating the LPHDR execution units' results according to whether repeating the same operation multiple times on the same input(s) yields an average result that differs from an exact result by at least a minimum specified percentage.

35. The "repeated execution" limitation has little to no import for the asserted patents' digital embodiments. This is because a digital system will generally be deterministic. As a result, even if the digital system yields an inexact result when executing a particular operation, the result will always be the same for each additional execution of the operation. As Singular itself has alleged, an operation "repeatedly performed" by a deterministic system "on a given set of input[] signals will always yield the same output signal." First Amended Complaint (Dkt. No. 37), ¶ 94. Accordingly, in conventional digital embodiments, the statistical mean of the results from "repeated execution" will be the same as the result of a single execution.

36. The import of the claim term "repeated execution" is instead tethered to the asserted patents' analog and hybrid analog-digital embodiments. As noted above, each additional execution of an operation in analog and hybrid analog-digital systems will not yield repeatable results because of noise. Consequently, the average result of an operation in an analog or hybrid analog-digital system will change as the same operation is repeatedly executed. These different results would have a statistical mean that would differ from the result of a single execution.

37. Because results are not repeatable in analog or hybrid analog-digital systems, the claim term "repeated execution" fails to delineate for those of skill in the art a method for measuring the difference between LPHDR execution units' results and the "exact" results of the same operation. Put another way, "repeated execution" of an operation will not reliably show

whether a device meets the claimed levels of difference because each execution can cause the difference between the average and exact result to change in an unpredictable way.

38. To illustrate the problem, I return to our earlier example of an analog system performing the operation "2 x 1." Let us assume that each execution of this operation yields a result that differs by up to ±5% from the exact result of 2—*i.e.*, the result of the operation can range from 1.9000 to 2.1000. Each execution of the operation has the potential to yield a different result within this range. As a result of the different possible results for each execution, the average result for all executions is prone to change with each execution. By extension, the degree to which the average result differs from the exact result of 2 is prone to change with each execution of the operation.

39. For example, consider the following ten hypothetical executions of the operation "2 x 1," which are provided purely for illustrative purposes:

Execution Number	Result of Execution	% Difference Between Result of Execution and Exact Result
1	2.0300	1.5000%
2	2.0105	0.5250%
3	1.9620	(1.9000%)
4	2.0013	0.0650%
5	2.0463	2.3150%
6	1.9506	(2.4700%)
7	2.0903	4.5150%
8	1.9163	(4.1850%)
9	1.9984	(0.0800%)
10	1.9960	(0.2000%)

40. In this scenario, each of the ten executions of the operation separately yields a result that differs from an exact result by well over ±0.05%, which is the degree of difference claimed in independent claim 1 of the '273 patent.

41. But consider the following table, which treats the execution results above as cumulative—for example, the third row shows the cumulative average result for the first three executions. As the table below shows, the delta between the exact result and the average result after each execution does not consistently reach the $\pm$0.05% threshold:

Number of Times Operation Is Executed	Cumulative Average Result of All Executions	% Difference Between Cumulative Average Result of All Executions and Exact Result
1	2.0300	1.5000%
2	2.0203	1.0125%
3	2.0008	0.0417%
4	2.0010	0.0475%
5	2.0100	0.5010%
6	2.0001	0.0058%
7	2.0130	0.6500%
8	2.0009	0.0456%
9	2.0006	0.0317%
10	2.0002	0.0085%

42. As this second table shows, were the operation executed two, five, or seven times, the average result would differ from an exact result by at least the claimed percentage of 0.05%. But were the operation executed three, four, six, eight, nine, or ten times, the system would yield an average result that differs from an exact result by less than the claimed percentage of 0.05%. All of those average results are the product of "repeated executions." Yet some differ from an exact result by at least the claimed percentage, and others do not.

43. In short, each potential execution would be a roll of the dice, creating the potential for an average result that might or might not show a possibility of infringement. There is no objective number of repeated executions that would determine whether a device infringes or not —instead, random chance would dictate whether a particular number of executions yielded an average result that differed from an exact result by a claimed percentage.

44. I see nothing in the claim language, specification, or prosecution history of the asserted patents that purports to resolve this dilemma. In fact, the only portions of the specification that discuss a specific application of repeated operations are irrelevant to the dilemma of unpredictable results generated by repeated execution of operations in analog and hybrid analog-digital systems.

45. Specifically, these portions of the specification discuss:

- a "nearest neighbor problem," which involves "a large set of vectors, called Examples, and a given vector, called Test," and the goal of "find[ing] the Example which is closest to Test where the distance metric is the square of the Euclidean distance (sum of squares of distances between respective components)"; and
- the "related" task of "Distance Weighted Scoring," in which "each Example has an associated Score" that "is a number that in some way characterizes the Example," and the goal, "[g]iven a Test Vector, . . . is to form a weighted sum of the Scores of all the Examples, where the weights are a diminishing function of the distances from the Test to the respective examples."

'273 patent at 17:33-37, 21:38-47.

46. The specification goes on to describe the results of several rounds of testing different variants of the algorithm, with each set of results corresponding to a set of ten runs comprising 100 tests each. '273 patent at 19:21-20:60.

47. For the nearest neighbor problem, the specification reports the percentage of "matches" from the tests. The "matches" refer to the percentage of times the algorithm correctly found the "nearest neighbor"—*i.e.*, the "Example closest to the Test." '273 patent at 19:30-33. In

other words, the figure describes only how often algorithm correctly identified the nearest neighbor—not the extent to which average results of repeated executions of the same operation differed from exact results.

48. As for the distance weighted scoring, the specification lists the various simulations' "mean score error." The "mean score error" refers to how well the algorithm approximated "weighted scores," which are computed "along with nearest neighbors." '273 patent at 19:59-60, 21:55-56. But while the specification states that "computing an overall weighted score involves summing the individual weighted scores associated with each example," '273 patent at 22:6-8, the summation in question was done through a method known as Kahan summation, '273 patent at 22:13-15, which is a method of ***reducing*** error in the summation of values to get closer to an exact result. The claims, by contrast, require a "statistical mean" of results that ***exceeds*** an exact result by a specified percentage.

49. In any case, even if the test results purported to describe comparisons between the average result of repeated execution of an operation and the exact result of that operation, it would not provide reasonable certainty about the boundaries of the invention. An analog or hybrid analog-digital embodiment is inherently unpredictable, meaning the results of executing analog and hybrid analog-digital operations are not repeatable.

50. Although one of skill in the art would know many methods to mitigate the problem of noise rendering analog operations non-repeatable, noise cannot be eliminated. No matter how much noise-generated differences in results can be reduced, noise will still be present and prevent any analog operation from being repeatable. This unpredictability of the results of any given execution of an operation is what makes the average result of repeated executions a moving target.

51. In sum, it is my opinion that the phrase "repeated execution" is indefinite because the asserted patents' claims, specification, and prosecution history do not give a skilled artisan reasonable certainty as to the number of repeated executions of an operation that are considered when calculating the "statistical mean" of resulting values that is to be compared with an exact result of the same operation.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct. Executed on January 9, 2021 at Seoul, Republic of Korea.

Gu-Yeon Wei

CERTIFICATE OF SERVICE

I certify that this document is being filed through the Court's electronic filing system, which serves counsel for other parties who are registered participants as identified on the Notice of Electronic Filing (NEF). Any counsel for other parties who are not registered participants are being served by first class mail on the date of electronic filing.

/s/ Nathan R. Speed
Nathan R. Speed